



National Board of Medical Examiners and Curriculum Change: What Do Scores Tell Us? A Case Study at the University of Balamand Medical School

Mode Al Ojaimi¹ , Megan Khairallah² , Rayya Younes², Sara Salloum² and Ghania Zgheib²

¹Faculty of Medicine, University of Balamand, Balamand, Lebanon. ²Department of Education, University of Balamand, Balamand, Lebanon.

Journal of Medical Education and Curricular Development
Volume 7: 1–9
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2382120520925062



ABSTRACT

OBJECTIVES: This study describes the results of NBME (National Board of Medical Examiners) implementation in Balamand Medical School (BMS) from 2015 to 2019, after major curricular changes were introduced as of 2012. BMS students' performance was compared with the international USMLE step 1 (United States Medical Licensing Examination, herein referred to as step 1) cohorts' performances. The BMS students' NBME results were analyzed over the successive academic years to assess the impact of the serial curricular changes that were implemented.

METHODS: This longitudinal study describes the performance of BMS preclinical second year medicine (Med II) students on all their NBME exams over 4 academic years starting 2015-2016 to 2018-2019. These scores were compared with the step 1 comparison group scores using item difficulty. The *t* test was computed for each of the NBME exams to check whether the scores' differences were significant.

RESULTS: Results revealed that all BMS cohorts scored lower than the international USMLE step 1 comparison cohorts in all disciplines across the 4 academic years except Psychiatry. However, the results were progressively approaching step 1 results, and the difference between step 1 scores and BMS students' NBME scores became closer and not significant as of year 4.

CONCLUSIONS: The results of the study are promising. They show that the serial curricular changes enabled BMS Med II students' scores to reach the international cohorts' scores after 4 academic years. Moreover, the absence of statistical difference between cohort 4 scores and step 1 cohorts is not module dependent and applies to all clinical modules. Further studies should be conducted to assess whether the results obtained for cohort 4 can be maintained.

KEYWORDS: National Board of Medical Examiners, medicine II curricular changes, assessment of analytical reasoning skills, multiple-choice questions, USMLE step 1 results

RECEIVED: April 15, 2020. **ACCEPTED:** April 16, 2020.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Megan Khairallah, Department of Education, University of Balamand, P.O. Box 33 Amioun, Balamand, Al-Koura, Lebanon. Email: megan.khairallah@balamand.edu.lb

Introduction

Medical education has evolved over the years from “didactic lectures, with each individual course delivering information on a sub-set of individualized topics related to the specific department where the course was instructed”^{1(p2)} toward more integration between the basic and clinical science courses. The purpose of this shift is to develop medical students' understandings of the residency requirements and their aptitude in medicine,^{1,2} on the one hand, and their analytical reasoning^{2,3} and critical thinking skills,⁴ on the other. These major curriculum changes are challenging and impact all aspects of the curriculum spanning from teaching methodology to assessment techniques.^{3,5-7}

Medical schools are adapting their curriculum and redefining the fundamental skills and abilities by including clinical reasoning skills and implementing new teaching methodologies that enhance critical thinking in the medical practitioner.⁴ Medical education is currently changing to integrate clinical skills as early as first preclinical year of Med I.^{1,8} Similarly,

some curricular changes focus on how didactic instruction can be combined or underpinned with case-based learning, problem-based learning, team-based learning, active learning, and independent learning exercises.¹ The value of illustrating teaching points with actual cases or simulated cases known as case-based learning is well recognized.⁹ Such integration of clinical and basic science courses has provided students with a broader knowledge with which they can accurately diagnose and treat patients. Curricular change would involve new teaching methodologies as well as appropriate assessment practices that support and measure the new desired student outcomes.¹⁰

Background

Changes to the curriculum

The integration of clinical reasoning into the curriculum has improved the quality of medical education and helped the students transition from medical school into the workplace.^{1,8}



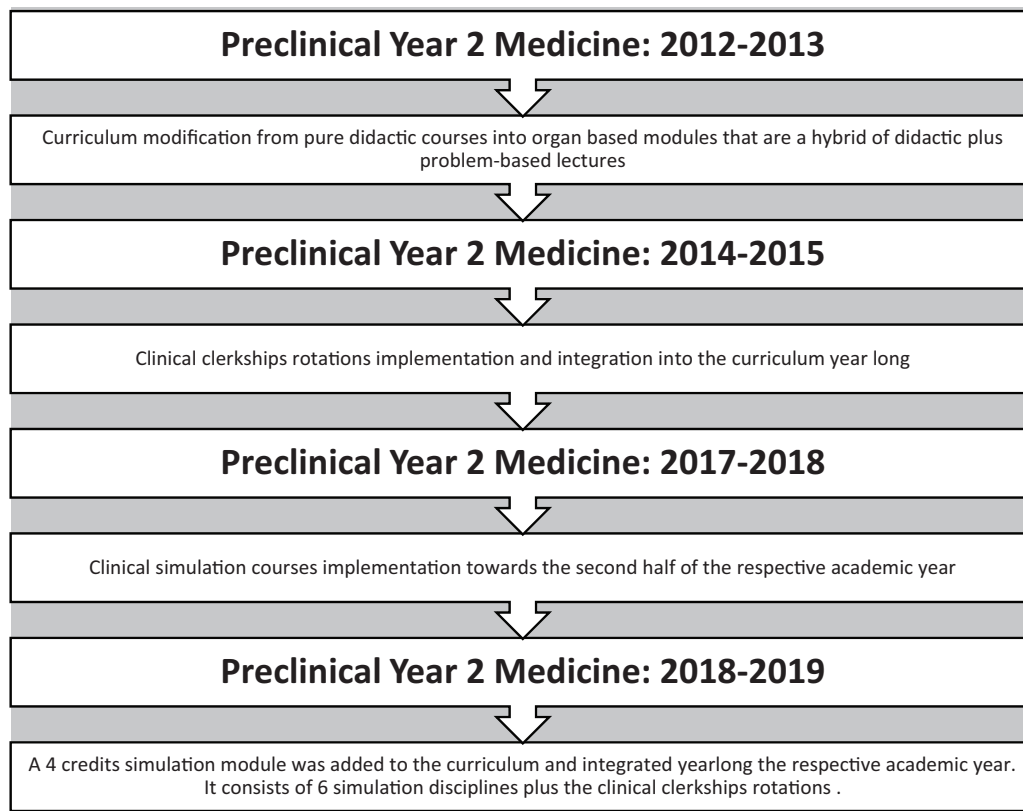


Figure 1. Med II curricular changes at BMS as of 2012.

In line with its pursuit of “excellence in medical education and patient care” and a commitment to a “dynamic curriculum,” Balamand Medical School (BMS) has been progressively revising its curriculum since 2012 to move away from more traditional “rote” learning and memory-based content approaches toward a hybrid curriculum with both didactic and problem and case-based learning approaches. While such revisions have been studied globally¹¹ and to a lesser extent in Lebanon such as in the American University of Beirut,¹² there is still vast room for research in the Lebanese and Middle Eastern contexts as well as in the area of measuring assessment practices.

At BMS, sequential changes were implemented in the Med II curriculum as of 2012 as outlined in Figure 1.

In 2012, the traditional subject-based (didactic-based courses) courses were modified or replaced by organ-based modules. The modified curriculum was designed to include organ-based modules, didactic lectures, and problem-based learning. In 2014-2015, a clinical skill’s module was integrated yearlong in Med II to introduce medical students to practical clinical experience early on and stimulate self-learning practices. In 2017-2018, simulation of Nursing Skills was implemented in Med II. In 2018-2019, more simulation courses were added to the Med II curriculum, such as Basic Life Support, Advanced Cardiac Life Support, Pharmacology of Cardiac Emergencies, Airway Management, and Ear and Prostate Examinations (normal and pathologic),

along with Ultrasound of Breast Pathologies and First Trimester Normal Pregnancy.

The BMS Med II hybrid curriculum was revised to ensure the optimization of clinical reasoning, evaluation, analysis, and inference among medical students as main competencies to be achieved. The desired outcomes from the curricular changes were that medical students become able to seek information, to integrate their learning with practical cases, and to critically and analytically reason about cases (ie, move away from memorization), as such outcomes serve as precursors to an accurate diagnosis.

Studies show that implementing simulations in a medical education helps the future practitioner develop communication and nontechnical and faster diagnosis skills, optimizing faster decision-making processes^{13,14}; it maximizes leadership and teamwork dynamics.¹³⁻¹⁵

Changes to assessment practices

In 2015, the Med II teaching committee chose to adopt the National Board of Medical Examiners (NBME) because it matched the hybrid nature of the revised curriculum in 2012 that focused on developing clinical analytical reasoning skills. As the curriculum changed, the nature of the exams needed to change as well.^{3,11,16,17} Assessment procedures had to reflect the shift from memorization to critical thinking and prepare students to a lifelong practice.¹⁸ It was with this

rationale that the NBME was adopted to monitor changes in performance and compare our curricula in light of international standards.¹⁹

Curricular changes necessitate using valid assessments both to align assessments with learning outcomes and to assess the extent and effectiveness of the changes.²⁰ Moreover, and as we discuss below, the NBME, as an international standardized assessment, helped raise standards at BMS and focus the curricular transformations toward the desired outcomes.

Prior to NBME implementation, multiple-choice questions (MCQs) generated by BMS faculty were used to assess Med II students. Even as MCQs can cover a large amount of content, there are growing concerns regarding the validity of the traditional institutionally generated multiple-choice examinations, as such types of exams are usually highly reliant on memorization and recall of knowledge rather than higher order thinking skills that require the learner to reach a conclusion, make a prediction, or select a course of action.^{18,21,22} The MCQs also have several weaknesses related to irrelevant difficulty and test-wiseness.²³ Correctly answering a question may not reflect the amount of knowledge about a tested topic and may simply relate to better test-taking strategies.²⁴

The NBME allows assessment of clinical reasoning, an essential lifelong skill that enables the medical practitioner to think and make decisions that guide differential diagnosis and plans of action²⁵ with patient safety implications. Teaching medical students clinical reasoning may avoid lethal mistakes.^{26,27} Clinical reasoning has become one of the most important competencies of health care practitioner's education and training, and it is highly recommended that this skill is taught to medical students as early as the first preclinical year.²⁶

The NBME questions focus on important concepts that are encountered in real clinical practice and assess the application of knowledge rather than recalling isolated facts. An "application of knowledge" item is defined as a question that requires the examinee to make a prediction, reach a conclusion, or select a course of action.^{18,28} These items require higher order skills compared with the recall items that test only examinees' knowledge of isolated facts without their application.²⁴ The NBME MCQs are of 2 types: true and false test items that require the students to select all the options that are true and 1 best answer item that requires the student to select 1 answer, and the latter is more widely used.²³ Examples of these items are "vignettes" that offer information in a table or graph format to assess the examiner's ability to reach the right conclusion and "problem-solving clinical vignettes" that assess the ability of the examinee to take the most appropriate next step. The NBME also offers the advantage to provide a detailed feedback to the students regarding their test performance promoting active learning strategies and guiding future learning. This allows the faculty and students to pinpoint areas of weakness and accordingly address them.

Aim of the study

Innovations in medical education methodology and evaluation differ based on contexts and cultures, and there is no one-size-fits-all curriculum. Medical schools are eager to embrace change, but those changes come at a price of researching these changes as new practices need to be based on reason, evidence, and best practices.^{3,10} Furthermore, there may be resistance to curricular changes on different levels,^{3,8,29} such as faculty lack of time to undergo the training in a busy and hectic environment.²⁹ Rapid changes in medical education are not easy to implement because "educational inertia—that is, the maintenance of the status quo—is a powerful force in most medical schools. Implementing successful educational programs across settings takes time, hard work, and financial resources."^{3(p1487)} Furthermore, active educational leadership and rigorous research are needed to successfully implement these innovations.^{3,6}

The aim of this study was to describe the results of NBME implementation in Balamand Medical School from 2015 to 2019. BMS students' performance was compared to international cohorts' performances. We also analyzed the BMS students' results over successive academic years to assess the impact of the serial curricular changes that were implemented.

The NBME as a standard assessment tool was implemented to position our students on an international scale as a way to validate our degrees at an international level. However, the exam then became a way to adjust, improve, and modify the curriculum to better reach the learning outcomes. It was hoped that the exam would serve as an external validation tool that would reflect the clinical reasoning potential of our Med II students and the progress of critical thinking with the newly integrated curriculum.

To validate the changes in the curriculum with reference to the NBME results as both a worldwide standard and an internal evaluation tool of our Med II students' performance, the following research questions were developed to guide the data analysis of the study:

1. How does the performance of BMS Med II students compare with an international cohort of Med II students on a similarly constructed NBME exam over a 4-year span?
2. Which modules reflect a wider gap in performance between BMS Med II students and the international scores? How can these gaps be explained?
3. What BMS curricular changes seem to impact BMS Med II students' performance?

Methods

The study was conducted over 4 academic years starting in 2015-2016 when NBME was implemented for the first time at

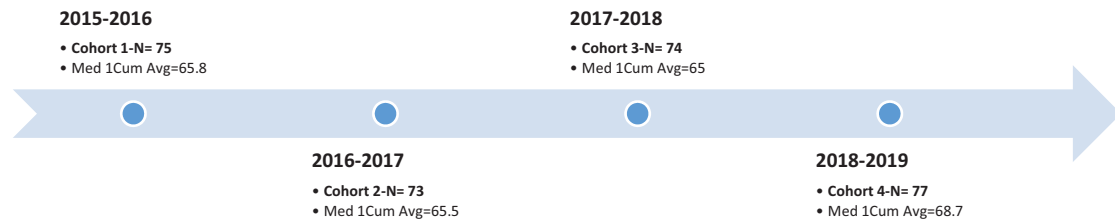


Figure 2. Cohorts' average across the years.

Table 1. ANOVA results of means of the students at the end of Med I for each of the 4 cohorts.

	SS	DEGREES OF FREEDOM	MEAN SQUARED	F	P VALUE
Between	629.36	3	209.79	3.65	.013
Within	17 054.4	297	57.42		
Total	17 683.8	300			

Abbreviations: ANOVA, analysis of variance; SS, sum of squares.

BMS. The University of Balamand Institutional Review Board approved the study. A total of 299 Med II students took the NBME exams at BMS. The students were clustered into 4 cohorts, and each cohort represents the Med II students in 1 academic year as seen in Figure 2. Figure 2 represents the number of students in each cohort (academic year) and the cohort yearly cumulative average for Med I, the year before the students started Med II.

The NBME exams were customized examinations constructed from the NBME bank questions by the clinical modules attended and reviewed by the respective module coordinators. These exams were administered at the end of each respective clinical module. No internal assessments for the respective modules other than the NBME were administered for these students during Med II. To examine BMS students' performance, students' scores on the different NBME tests were compared with the USMLE step 1 comparison group scores of those who took the same questions using the test item difficulty. Item difficulty represents the percentage of students who responded correctly to the item. The test item difficulty is the average item difficulty of all items in the test. These data are provided by the detailed report issued by the NBME to BMS at the end of each exam.

Med I yearly cumulative averages of each respective cohort were compared using 1-way 4-level analysis of variance (ANOVA) to check whether the students' performances were similar before entering Med II. The cohorts' cumulative average was based on each student's average on all the tests taken in Med I.

Then, for every academic year, a *t* test was calculated for each NBME test the students took to check whether the difference in scores between our students and the USMLE step 1 comparison group was significant.

In addition, a comparative study was performed comparing Med II students' performance over the span of the 4

listed academic years with step 1 cohort results in the same specific years.

Results

First, we looked at the performance of each cohort before the start of the year. A 1-way 4-level between-subjects ANOVA was conducted to compare the means of the students at the end of Med I for each of the 4 cohorts. The difference between the means was statistically significant, $P = .013$ (see Table 1). Because the results were statistically significant, we added Tukey post hoc tests to check where the differences are.³⁰

The post hoc tests compare each cohort mean with the other cohort means. The results of the post hoc tests are presented in Table 2. The Tukey post hoc tests indicated that the means of cohorts 1, 2, and 3 are not statistically significantly different from each other. In addition, the means of cohorts 1 and 4 are not statistically significantly different. However, the means of cohorts 2 and 3 are statistically significantly different from the mean of cohort 4. These results suggest that before starting Med II, the performance level of cohorts 1, 2, and 3 was similar. However, cohort 4 outperformed cohorts 2 and 3.

Next, we compared each cohort to the step 1 comparison group using the NBME tests. During 2015-2016, cohort 1 scored lower than the step 1 comparison group in all tests, as depicted in Table 3. This difference was statistically significant ($P < .05$) in 60% of the modules, namely, Pulmonary, Cardiovascular, Nephrology-Urology, Neurology, Dermatology, and Endocrine and Reproductive.

During 2016-2017, cohort 2 scored lower than step 1 cohort in all modules except Hematology-Oncology. This difference was statistically significant only in 40% of the modules: Nephrology-Urology, Pulmonary, Neurology, and Cardiovascular modules. The scores were matching for the 2 groups in Hematology-Oncology, as depicted in Table 4.

Table 2. Results of the post hoc tests on the means of the 4 cohorts.

	DIFFERENCE	95% CONFIDENCE INTERVAL		P VALUE
		LOWER	UPPER	
		-3.49	2.89	.99
Cohort 3–Cohort 1		-4.0	2.4	.92
Cohort 4–Cohort 1	Cohort 2–Cohort 1	-0.28	6.08	.09
Cohort 3–Cohort 2	-0.5	-3.71	2.71	.98
Cohort 4–Cohort 2	3.2	0.01	6.39	.048
Cohort 4–Cohort 3	3.7	0.50	6.90	.016

Table 3. Comparison of BMS cohort 1 scores (2015-2016) in different exam modules with step 1 cohort who took the same questions.

MODULE NAME	NO. OF QUESTIONS	COHORT 1 SCORES, %	STEP 1 SCORES, %	T TEST
Dermatology	60	63	76	<.001
Pulmonary	134	71	81	<.001
Nephrology	100	67	76	.023
Neurology	123	68	76	<.001
Cardiovascular	119	73	81	<.001
Endocrinology and Reproductive	123	74	80	.002
Hematology-Oncology	107	73	77	.095
Gastroenterology	126	74	78	.290
Basic Pharmacology	52	74	78	.109
Psychiatry	121	76	80	.090

The bold face values are statistically significant.

During 2017-2018, cohort 3 scored statistically significantly lower than step 1 cohort in all modules except Psychiatry where the score was lower than step 1 cohort; however, the difference was not significant, as shown in Table 5.

During 2018-2019, cohort 4 scores were lower than the step 1 cohort scores in all modules except Psychiatry, Nephrology, and Gastroenterology; however, the difference was not statistically significant in any of the modules, as shown in Table 6.

When comparing BMS students' overall performance in NBME since its implementation in 2015 until 2019, we notice that it is getting progressively closer to step 1, especially cohort 4, as shown in Graph 1. Cohort 3 values, however, appear to be dropping.

Discussion

While new curriculum and new testing are both time-consuming and costly, research on the NBME subject exam in medicine has been used to both measure the effects of new programs and compare students' knowledge-based performances.²⁸ Our results show that all BMS cohorts scored

lower than NBME cohorts in all disciplines except Psychiatry. The difference was significantly lower for cohorts 1, 2 and 3. This difference was not statistically significant for cohort 4. The lower performance in the Psychiatry module was statistically significant in cohort 1 only. It is worth noting that this performance was higher than step 1 cohort in 2018-2019, although the difference is not statistically significant.

The results show that it took 4 years after NBME implementation for the difference in the NBME scores between BMS and the comparative USMLE cohort to become statistically not significant. This lag period seems logical and reflects the required time for the students and faculty to readjust and adapt to this new assessment tool. It is worth mentioning that cohort 4 was introduced to NBME exams in Med I. This may have impacted the results of Med II NBMEs positively. Moreover, cohort 4 attended more simulation courses than any other cohort, which may have contributed to their enhanced performance compared with the step 1 comparison group in that year.

Table 4. Comparison of BMS cohort 2 (2016-2017) scores in different exam modules with step 1 cohort who took the same questions.

MODULE NAME	NO. OF QUESTIONS	COHORT 2 SCORES, %	STEP 1 SCORES	T TEST
Cardiovascular	123	70	81	<.001
Neurology	125	68	78	<.001
Pulmonary	134	72	81	<.001
Nephrology	123	69	76	.002
Basic Pharmacology	64	71	75	.194
Gastroenterology	124	74	78	.123
Immunology	50	81	84	.36
Endocrinology and Reproductive	124	79	81	.533
Psychiatry	121	79	80	.622
Hematology-Oncology	101	78	78	.705

The bold face values are statistically significant.

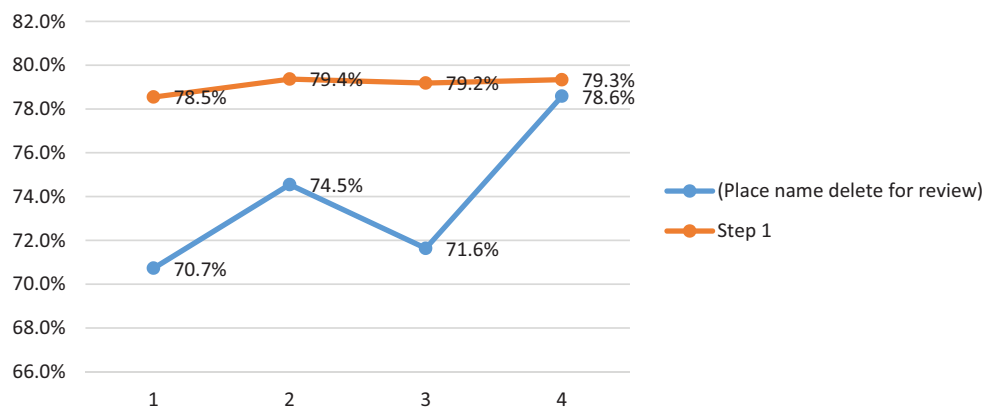
Table 5. Comparison of BMS cohort 3 (2017-2018) scores in different exam modules with step 1 cohort who took the same questions.

MODULE	NO. OF QUESTIONS	COHORT 3 SCORES, %	STEP 1 SCORES, %	T TEST
Nephrology	125	74	79	0.029
Pulmonary	134	69	81	<0.001
Cardiovascular	121	68	79	<00.001
Basic Pharmacology	64	69	75	0.026
Endocrinology and Reproductive	130	74	79	0.030
Psychiatry	121	77	80	0.172
Neurology	125	68	77	<0.001
Gastroenterology	124	71	78	0.005
Hematology- Oncology	120	73	78	0.023
Immunology	50	78	84	0.049
Dermatology	50	67	81	0.001

The bold face values are statistically significant.

Table 6. Comparison of BMS cohort 4 (2018-2019) scores in different exam modules with step 1 cohort who took the same questions.

MODULE	NO. OF QUESTIONS	COHORT 4 SCORES, %	STEP 1 SCORES, %	T TEST
Basic Pharmacology	108	73	74	0.789
Cardiovascular	125	78	80	0.45
Pulmonary	125	78	79	0.447
Endocrinology and Reproductive	126	79	80	0.662
Dermatology	50	71	78	0.105
Neurology	125	78	79	0.815
Psychiatry	95	84	82	0.243
Nephrology	125	79	77	0.38
Gastroenterology	125	79	79	0.93
Immunology	50	83	84	0.822
Hematology-Oncology	120	83	80	0.079



Graph 1. Overall BMS Med II students' performance in NBME compared with that of step 1 cohorts from 2015 to 2016 (cohort 1), 2016 to 2017 (cohort 2), 2017 to 2018 (cohort 3), and 2018 to 2019 (cohort 4) (see Appendix 1 for Excel sheet to accompany Graph 1). NBME indicates National Board of Medical Examiners.

Medical schools are increasingly using simulation-based education to develop the medical students' cognitive abilities and interpersonal skills.¹⁴ Accordingly, a growing body of research is examining how the use of simulations impacts both the teaching methodology and the design of assessments in medical education.³¹ Integrating simulations into the medical curriculum has been found to improve performance on assessments³¹ by enabling medical students to integrate their knowledge with their clinical reasoning skills. This leads to an improvement in the learning curve and better knowledge retention and decision-making, especially in crisis.¹⁴

After 4 years of NBME implementation, BMS average was statistically similar to USMLE step 1 cohorts' averages, and these results have numerous implications and can be attributed to both academic (curricular and pedagogical) and student factors. Pedagogically, course instructors have modified their teaching methodologies and lectures to encourage more active participation and critical analysis. These factors could have been potential reasons for the improved performance in cohort 4 and are worth studying in more detail. The students themselves over these 4 years became more familiar with NBME exams. Each class may have been able to give better advice about the NBME test-taking skills and best resources to study the material to pass on to their junior cohorts. Examinees who tested in earlier cohorts may memorize important clues about the exams and provide them to subsequent cohorts (even though such conduct is prohibited). This diffusion might have happened at a superficial level when, for example, an examinee later tells a friend something.³¹ Moreover, fear of the new NBME may have lessened progressively each year as students realized that this type of assessment was part of the curriculum and "here to stay."

The performance of cohort 3 on NBME was suboptimal compared with the other 3 cohorts. Although this was not investigated in this study, we believe it is secondary to factors related to the students in this cohort. This may be an indirect indication that the improved performance across time is mainly due to learning and not to students getting used to taking standardized tests.

Previous research in the Lebanese context indicates that changing teaching practices at the university level faces different challenges such as faculty's concerns that covering the required material will be jeopardized, their beliefs about what represents good and effective teaching in the sciences and medical school, and the limited professional preparation that faculty receive.³² Moreover, literature has proposed that in addition to the aforementioned challenges, tension may exist between the professional identity of scientists, science professors, and pedagogical reform, and that the challenge of addressing such tensions remains a complicated issue.³³ In this study, adopting the NBME as a change in assessment practices and the other curricular changes such as incorporation of simulations seemed to motivate instructors to change their practices. Therefore, curricular change needs to be gradual, providing sufficient time for faculty to adopt it,³⁴ and importantly, curricular changes need to be aligned with valid assessments that monitor the desired outcomes. Utilizing assessments like the NBME provides a mechanism for faculty to receive constant feedback on students' progress toward desired and valuable outcomes. Our study shows that when the faculty is able to note relations between changing practices and students' outcome, they are in a better position to address the challenges identified earlier.³⁵ Results of this research and our students' recent results suggest that actively aligning curriculum changes with assessment practices facilitates curricular reforms by pinpointing the misalignments and pushing changes to reach the desired and previously missed learning outcomes.

One limitation is the longitudinal time period over which this study took place. While the difference between BMS Med II cohort 4 scores and NBME is not significant, further results from forthcoming academic years are not available yet to show that these results are sustainable. Further research should extend to continue measuring the impact of the curricular changes. Another limitation is that this study does not address the factors which could have contributed to the lower performance observed in cohort 3.

Conclusion

Implementing NBME in BMS was a challenge for both the students and faculty. The results of our study are promising, though, as they demonstrate that the progressive curricular changes enabled BMS Med II students' scores to catch up with the international cohorts after 4 academic years. Moreover, the absence of statistical difference between cohort 4 scores and step 1 cohorts is not module dependent and applies to all clinical modules.

Further studies should be conducted to assess whether the results obtained for cohort 4 can be maintained in future years. It is worth studying the scores of these students in step 1 and correlating them with their Med II NBME scores.

Do NBME scores indicate curricular success or does NBME instigate curricular changes?

Sustainable long-term change requires assessments that can first support *and* later help evaluate the effects of the changes for continuous improvements. Our study shows that curricular changes should be supported by changes in assessment practices, and assessment results may take longer to show a clear indication of improvement. Our results should encourage other medical schools undergoing curricular changes to look at long-term goals. Curricular changes need modifications in teaching methodology and in content per module, as well as changes in assessment practices. While this study focuses on assessment as a successful indicator of change, the other changes in the curriculum need equal exploration. For better validity, more NBME implementation data from other international and Lebanese medical schools using similar measurement must be further investigated.

Acknowledgements

The authors gratefully acknowledge the invaluable contributions of Ms. Rania Najjar - Senior Data Analyst, University of Balamand - for her support in the data analysis.

Author Contributions

All the authors were involved in conception, literature review, drafting and revising of the manuscript. All the authors approved the final manuscript.

ORCID iDs

Mode Al Ojaimi  <https://orcid.org/0000-0001-8256-5395>

Megan Khairallah  <https://orcid.org/0000-0001-7231-0989>

REFERENCES

- Myers S. The changing face of medical education [published online ahead of print March 27, 2014]. *J Med Educ Curric Dev*. doi:10.4137/JMECD.S17270.
- Bailey J, Perkowski L. Combining expertise: reflecting on a team approach to curriculum development and implementation. *J Med Educ Curric Dev*. 2016;3:25-31. doi:10.4137/JMECD.S38104.
- McGaghie W, Barsuk J, Cohen E, Kristopaitis T, Wayne D. Dissemination of an innovative mastery learning curriculum grounded in implementation science principles. *Acad Med*. 2015;90:1487-1494. doi:10.1097/acm.0000000000000907.
- Morrissey B, Heilbrun ME. Teaching critical thinking in graduate medical education: lessons learned in diagnostic radiology [published online ahead of print March 20, 2017]. *J Med Educ Curric Dev*. doi:10.1177/2382120517696498.
- Jarvis-Selinger S, Hubinette M. The matrix: moving from principles to pragmatics in medical school curriculum renewal. *Acad Med*. 2018;93:1464-1471. doi:10.1097/acm.0000000000002306.
- Velthuis F, Varpio L, Helmich E, Dekker H, Jaarsma A. Navigating the complexities of undergraduate medical curriculum change. *Acad Med*. 2018;93:1503-1510. doi:10.1097/acm.0000000000002165.
- Johnson T, Khalil M, Pepler R, Davey D, Kibble J. Use of the NBME comprehensive basic science examination as a progress test in the preclerkship curriculum of a new medical school. *Adv Physiol Educ*. 2014;38:315-320. doi:10.1152/advan.00047.2014.
- Quintero G, Vergel J, Arredondo M, Ariza M, Gomez P, Pinzon-Barrios AM. Integrated medical curriculum: advantages and disadvantages. *J Med Educ Curric Dev*. 2016;3:133-137. doi:10.4137/jmeed.S18920.
- McLean S. Case-based learning and its application in medical and health-care fields: a review of worldwide literature. *J Med Educ Curric Dev*. 2016;3:339-349. doi:10.4137/jmeed.S20377.
- Halalau A, Falatko J, Mi M. Application of adult learning theory in teaching evidence based medicine to residents. *J Med Educ*. 2016;15:185-193.
- Gullo C, Dzwonek B, Miller B. A disease-based approach to the vertical and horizontal integration of a medical curriculum. *Med Sci Educ*. 2016;26:93-103. doi:10.1007/s40670-015-0208-5.
- Zgheib N, Simaan J, Sabra R. Using team-based learning to teach pharmacology to second year medical students improves student performance. *Med Teach*. 2010;32:130-135. doi:10.3109/01421590903548521.
- Jakobsen R, Gran S, Grimsø B, et al. Examining participant perceptions of an interprofessional simulation-based trauma team training for medical and nursing students. *J Interprof Care*. 2018;32:80-88. doi:10.1080/13561820.2017.1376625.
- Reis S, Guimarães P, Coelho F, Nogueira E, Coelho L. A framework for simulation systems and technologies for medical training. Paper presented at: 2018 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE); March 19-24, 2018; Porto, Portugal.
- Noureldin Y, Lee J, McDougall E, Sweet R. Competency-based training and simulation: making a "valid" argument. *J Endowol*. 2018;32:84-93. doi:10.1089/end.2017.0650.
- Lundy M, Standley C, Westveld A. Curricular reform in two medical school tracks and the impact on USMLE scores. *Med Sci Educ*. 2017;27:201-207. doi:10.1007/s40670-016-0368-y.
- Ryan M, Bishop S, Browning J, et al. Are scores from NBME subject examinations valid measures of knowledge acquired during clinical clerkships? *Acad Med*. 2017;92:847-852. doi:10.1097/acm.0000000000001535.
- Vanderbilt A, Feldman M, Wood IK. Assessment in undergraduate medical education: a review of course exams. *Med Educ Online*. 2013;18:1-5. doi:10.3402/meo.v18i0.20438.
- Achampong E. Assessing the current curriculum of the nursing and midwifery informatics course at all nursing and midwifery institutions in Ghana [published online ahead of print January-December 2017]. *J Med Educ Curric Dev*. doi:10.1177/2382120517706890.
- Laverty J, Underwood S, Matz R, et al. Characterizing college science assessments: the three-dimensional learning assessment protocol. *PLoS ONE*. 2016;11:e0162333. doi:10.1371/journal.pone.0162333.
- Bauer D, Holzer M, Kopp V, Fischer M. Pick-N multiple choice-exams: a comparison of scoring algorithms. *Adv Health Sci Educ Theory Pract*. 2010;16:211-221. doi:10.1007/s10459-010-9256-1.
- Schuwirth L. How to write short cases for assessing problem-solving skills. *Med Teach*. 1999;21:144-150. doi:10.1080/01421599979761.
- Paniagua M, Swygert K. *Constructing Written Test Questions for the Basic and Clinical Sciences*. 4th ed. Philadelphia, PA: National Board of Medical Examiners; 2016.
- Case S, Swanson D. *Constructing Written Test Questions for the Basic and Clinical Sciences*. 2001-2002. https://www2.med.wayne.edu/elab/guides/exam_writing/NBME_Test_Question_Writing_Guide.pdf.
- Charlin B, Tardif J, Boshuizen H. Scripts and medical diagnostic knowledge. *Acad Med*. 2000;75:182-190. doi:10.1097/00001888-200002000-00020.
- Amey L, Donald KJ, Teodorczuk A. Teaching clinical reasoning to medical students. *Br J Hosp Med*. 2017;78:399-401. doi:10.12968/hmed.2017.78.7.399.
- Gay S, Bartlett M, McKinley R. Teaching clinical reasoning to medical students. *Clin Teac*. 2013;10:308-312. doi:10.1111/tet.12043.
- Torre D, Papp K, Elnicki M, Durning S. Clerkship directors' practices with respect to preparing students for and using the national board of medical examiners subject exam in medicine: results of a United States and Canadian survey. *Acad Med*. 2009;84:867-871. doi:10.1097/acm.0b013e3181a858ef.
- Damp J, Dewey C, Wells Q, Horn L, Kroop S, Mendes L. Faculty development on clinical teaching skills: an effective model for the busy clinician. *J Med Educ Curric Dev*. 2016;3:187-191. doi:10.4137/JMECD.S40798.

30. Thompson B. *Foundations of Behavioral Statistics: An Insight Based Approach*. New York, NY: The Guilford Press; 2008.
31. Short K, Bucak D, Rosenthal F, Raymond M. When listening is better than reading: performance gains on cardiac auscultation test questions. *Acad Med*. 2018;93:781-785. doi:10.1097/acm.0000000000001906.
32. El Khoury R, BouJaoude S, El-Hage F. Étude exploratoire des points de vue épistémologiques des enseignant(e)s universitaires de sciences: cas d'une faculté de sciences Libanaise et francophone. *Spiral- E —Revue de Recherches en Éducation*. 2014;53:41-65.
33. Brownell S, Tanner KD. Barriers to faculty pedagogical change: lack of training, time, incentives, and . . . tensions with professional identity? *CBE Life Sci Educ*. 2012;11:339-346.
34. Kennedy M. How does professional development improve teaching? *Rev Educ Res*. 2016;86:945-980.
35. Guskey T. Professional development and teacher change. *Teac Teac Theo Prac*. 2002;8:381-391. doi:10.1080/135406002100000512.

Appendix 1. Excel sheet to accompany Graph 1 (formatting purposes only).

COHORTS' LABELS	(NAME DELETED FOR REVIEW PURPOSES)	STEP 1
1	70.7%	78.5%
2	74.5%	79.4%
3	71.6%	79.2%
4	78.6%	79.3%
Grand total	0.739777778	0.791111111